

METABOLÔMICA E QUIMIOMETRIA COMO FERRAMENTAS PARA ANÁLISES QUIMIO(BIO) DIVERSAS

Marcelo Maraschin
Amélia R. Somensi Zeggio
Maíra M. Tomazzoli
Simone K. Oliveira
Fernanda Ramlov
Maria Beatriz Veleirinho
Miguel Rocha

1.1 INTRODUÇÃO

1.1.1 Metaboloma e métodos analíticos

No presente, a humanidade vive uma revolução cultural caracterizada pelos avanços tecnológicos e pelo intenso fluxo de informações, comparável em impacto ao surgimento da agricultura há aproximadamente 10 mil anos e à Revolução Industrial iniciada no século XVIII¹. Algumas inovações

tecnológicas ocorridas nas três últimas décadas levaram a uma nova forma de pensar os sistemas biológicos e, principalmente, de pesquisá-los^{2,3}. Na área biológica, o marco inicial desta revolução foi denominado de era genômica, sendo caracterizado pelo desenvolvimento, padronização e otimização das técnicas de engenharia genética⁴. Os ensaios genômicos evoluíram rapidamente, e o imenso volume de dados genéticos possibilitou o aprofundamento da análise das variações temporais e espaciais no acúmulo de transcritos, proteínas e metabólitos. Essa nova fase da atual revolução na área biológica foi denominada de era pós-genômica, ou da genômica funcional⁵.

O termo metaboloma foi cunhado em 1998, denominando a análise da concentração relativa de metabólitos resultante de modificações no padrão de expressão gênica⁶. A metabolômica tem como pressuposto uma abordagem holística e interativa de análise, segundo a qual o metabolismo celular é efetivamente entendido como uma rede complexa de reações altamente ordenadas e interconectadas, de modo que mesmo pequenas alterações, como a diminuição na concentração ou atividade de uma enzima, podem causar modificações simultâneas na concentração de centenas de metabólitos. Desse modo, o metaboloma fornece uma ligação direta entre o genoma, o transcriptoma e o proteoma, podendo revelar quais fatores influenciam diretamente uma dada função biológica⁷⁻⁹.

A originalidade do uso do termo metaboloma proposto por Oliver e colegas (1998)⁶ deve-se não a inovações nas técnicas analíticas utilizadas, mas sim ao estabelecimento do uso conjunto daquele ferramental de análise como complemento às outras áreas “ômicas”⁴. A partir de então, o termo passou a designar uma estratégia de análise bastante complexa, que se refere ao levantamento qualitativo e quantitativo dos metabólitos presentes em um organismo, ou em um dado componente deste (tecidos ou células, por exemplo), denominando-se metaboloma parcial neste último contexto^{6,10-12}.

Alterações induzidas geneticamente, epigeneticamente ou por influência do ambiente são manifestadas, em última instância, através de alterações na composição e concentração de metabólitos. Assim, comparando-se os perfis metabólicos gerados em tecidos que diferem geneticamente ou no seu estado epigenético diferenças genômicas funcionais podem ser inferidas^{2,4,10,13}.

Até o presente, a utilização de uma única técnica analítica não possibilita a obtenção do quadro metabolômico completo de uma dada amostra, sendo necessária a adoção de um conjunto interdisciplinar de abordagens que integra a biologia, a química analítica, a química orgânica, a quimiometria e a informática¹². As técnicas analíticas comumente utilizadas na metabolômica são as cromatografias líquida (CL) ou líquido-gasosa (CG), associadas ou não

à espectrometria de massas (EM) e as espectroscopias de infravermelho (espectroscopia vibracional de infravermelho médio, ou *Fourier-transform infrared spectroscopy* – FTIR; e espectroscopia vibracional de infravermelho próximo, ou *near infrared spectroscopy* – NIR) e de Raman, bem como a espectroscopia de ressonância magnética nuclear de hidrogênio (^1H -RMN), de carbono (^{13}C -RMN) e de nitrogênio (^{15}N -RMN), em suas abordagens experimentais uni (1D) e bidimensionais (2D) e também quantitativas, isto é, espectroscopia de ressonância magnética nuclear quantitativa – qRMN^{14,15,16,17}.

Os estudos metabolômicos podem ser conduzidos de modo seletivo ou de modo não seletivo. No caso das análises seletivas, a atenção focaliza-se na caracterização de um composto, ou de um conjunto destes, pertencente a uma classe específica de metabólitos (*target compounds*), permitindo um aumento na sensibilidade e, normalmente, preocupando-se com a quantificação do(s) analito(s). Técnicas de CL, CG e qRMN são normalmente utilizadas nessa forma de análise. De outra maneira, as análises conduzidas no modo não seletivo visam à observação simultânea do maior número de compostos nas amostras, podendo gerar perfis metabólicos (*metabolic profiles*) ou padrões metabólicos (*metabolic fingerprints*) característicos de um dado fenótipo. Técnicas como 1D e 2D-RMN, FTIR, NIR e EM normalmente são utilizadas na execução dessa abordagem analítica não seletiva^{8,11,18}. As características básicas das técnicas analíticas mais frequentemente utilizadas e suas aplicações na análise metabolômica são mostradas na Tabela 1.1.

Tabela 1.1 Técnicas analíticas de uso corrente, suas características e aplicações em estudos metabolômicos

TÉCNICA ANALÍTICA	CARACTERÍSTICAS	APLICAÇÕES
CG-EM	Alta sensibilidade, analitos voláteis somente, baixo custo	Metabólito(s) alvo(s) e metabolômica
CL-EM	Alta sensibilidade, analitos não voláteis/ termoinstáveis, alto custo	Metabólito(s) alvo(s), perfil metabólico e metabolômica
RMN	Baixa sensibilidade, alta reprodutibilidade, não destrutiva, alto custo	Perfil metabólico, padrão metabólico (<i>fingerprint</i>) e fluxo metabólico
EM	Alta sensibilidade, alto custo, usualmente acoplada à técnica de separação	Metabólito(s) alvo(s), perfil metabólico, padrão metabólico (<i>fingerprint</i>), metabolômica e fluxo metabólico
FTIR, NIR e Raman	Baixa resolução, alta reprodutibilidade e rendimento, não destrutiva, baixo custo	Padrão metabólico (<i>fingerprint</i>)

Após a geração dos dados metabolômicos, estes são geralmente tratados através de métodos de análise multivariada, em função do enorme volume

de informações que deve ser considerado (32 mil dados/espectro ^1H -RMN, por exemplo) e também porque grande parte do que se deseja elucidar são relações complexas, lineares ou não lineares. O uso da matemática e de ferramentas da estatística à análise de um (mega)conjunto de informações químicas recebe a denominação de quimiometria. Os métodos de análise multivariada normalmente utilizados na quimiometria incluem a análise dos componentes principais (PCA), a análise de agrupamentos hierárquicos (AH), a regressão múltipla e as análises discriminatória, fatorial e canônica^{12,19}. O uso combinado dessas ferramentas, que serão descritas mais detalhadamente na Seção 1.4, permite a extração de conhecimento útil a partir dos dados em bruto, incluindo a identificação dos metabólitos nas amostras, os metabólitos cuja concentração varia mais significativamente entre grupos de amostras, o agrupamento de amostras e/ou metabólitos por similaridade, ou ainda a classificação das amostras em grupos consoante aos dados metabolômicos.

1.2 HISTÓRICO

A partir do final do século XIX, a química analítica tem experimentado um intenso processo de evolução, levando à descoberta e ao desenvolvimento de novas moléculas com uma ampla gama de aplicações industriais. Para tal, um contínuo desenvolvimento de técnicas e métodos analíticos tem sido observado, buscando otimizar o desempenho destes através da geração de resultados com maior confiabilidade, rapidez e baixo custo.

Em muitos casos, a análise química é utilizada em estudos bioquímicos (química bioanalítica ou bioanálise), onde a quantidade do analito de interesse é bastante pequena, de modo que o uso de técnicas de grande seletividade e sensibilidade é de fundamental importância. De fato, em bioanálise há uma contínua demanda à determinação de compostos em matrizes complexas, isto é, biofluidos, cujos níveis de concentração situam-se, não raro, na faixa de pico/fentomol ($10^{-12}/10^{-15}$ M). Este aspecto é especialmente importante, por exemplo, no desenvolvimento de farmoquímicos mais potentes, os quais podem ser administrados em concentrações mais baixas²⁰.

Em estudos metabolômicos, é frequente o uso de material amostral de alta complexidade química, a exemplo do observado em extratos de tecidos vegetais e animais, de micro-organismos e biofluidos. Assim, é mister considerar que o efeito da complexidade da matriz amostral poderá gerar um conjunto de dados com baixa precisão em alguns casos. A estratégia usual

para superar essa dificuldade baseia-se na utilização de métodos analíticos sequenciais em bioanálise, de modo a reduzir o efeito de fontes de perturbações (*noise*) oriundas da matriz e/ou do instrumental. A abordagem mais simples nesse contexto considera, inicialmente, a adoção de procedimentos de pré-tratamento amostral (*clean-up*), buscando isolar ou pré-concentrar o(s) composto(s) de interesse, contribuindo à seletividade do processo dessa forma. Para a remoção de compostos de alto peso molecular (proteínas, ácidos nucleicos, polissacarídeos, por exemplo) e enzimas, a desnaturação, a ultrafiltração, a diálise e a permeação em gel poderão ser utilizadas. Além disso, a extração em sistemas líquido-líquido, ou em fase sólida, poderá ser útil à remoção de sais, ou ainda como etapa final no processo de purificação do composto. A despeito do número de etapas de pré-tratamento da amostra, é importante considerar nessa fase os riscos de contaminação, a adsorção irreversível dos analitos à matriz de separação, a decomposição da amostra e a não reprodutibilidade de alguma etapa no processo analítico sequencial^{20,21}. Adicionalmente, enfatiza-se a necessidade de padronização de métodos de pré-tratamento amostral em análise metabolômica, especialmente em estudos intra/interlaboratorial de natureza comparativa e de estabelecimento de padrões metabólicos²². Assim, independentemente da técnica analítica subsequente a utilizar, isto é, seletiva ou não seletiva de analitos, a precisão observada no conjunto de dados metabolômicos mostra-se como função direta de protocolos padronizados de preparo amostral.

Após a etapa de preparo amostral (se necessário), considera-se a utilização de técnicas seletivas e de boa resolução, por exemplo, CL e CG, usualmente acopladas à EM, na análise quali/quantitativa de metabólitos de forma adequada aos objetivos de estudos metabolômicos focados em compostos alvos (*target metabolites* – Tabela 1.1). Essa realidade tornou-se possível a partir das décadas de 1950 e 1960, com o lançamento dos primeiros cromatógrafos gasosos, e expandiu-se nos anos 1980, com a adoção de colunas capilares e o acoplamento à espectrometria de massas, possibilitando a identificação simultânea de centenas de compostos em matrizes complexas²³⁻²⁷. De forma similar, a análise de moléculas não voláteis e termoinstáveis avançou mais significativamente com o desenvolvimento da CL, notadamente a partir dos anos 1980²⁰. Em função do exposto, quando a análise metabolômica busca identificar compostos-alvos, observa-se a necessidade de utilização de técnicas que sejam simultaneamente seletivas e com alta sensibilidade, possibilitando a obtenção de dados robustos na análise de amostras biológicas. Infelizmente, os sistemas de detecção utilizados isoladamente não oferecem, na maioria das vezes, a seletividade necessária, de

modo que a detecção do composto de interesse pode ser dificultada pela interferência de outras moléculas de alto ou baixo peso molecular, com natureza hidrofóbica ou hidrofílica, as quais ocorrem frequentemente em altas concentrações na matriz da amostra²⁸.

Uma segunda abordagem experimental em análises do metaboloma assume como pressuposto a não seletividade de alvos metabólicos em matrizes complexas de interesse. A obtenção de perfis metabólicos (*metabolic profiles*) associada ao estudo da função dos metabólitos^{27,29,30} encontra nas técnicas de EM e RMN suas principais ferramentas (Tabela 1.1).

1.3 POSSIBILIDADES TERAPÊUTICAS E INDUSTRIAIS

Nos últimos anos, é possível observar o crescimento contínuo das aplicações da metabolômica em diversas áreas de importância tecnológica. Por exemplo, um conjunto de métodos tem sido estabelecido para a determinação do metaboloma diferencial de alimentos (ou matérias-primas) transgênicos e não transgênicos^{31,32}, para a identificação de marcadores bioquímicos de tumores³³⁻³⁵, bem como para o estudo da diversidade biológica de espécies de interesse medicinal³⁶⁻³⁸ e alimentar^{18,39-41}. Nesse contexto, na sequência utilizar-se-á como modelo de estudos a análise metabolômica da própolis catarinense, a partir de um megaconjunto de dados espectrais de 1D e 2D-RMN e ferramentas de bioinformática. Um resumo esquemático do modelo de estudos proposto é descrito na Figura 1.1.

A análise metabolômica baseada na aplicação de métodos de reconhecimento de padrões de dados espectrais de 1D-RMN de referência (*metabolic fingerprint*) ou compostos alvo (*target metabolites*) a amostras de extratos hidroalcoólicos de própolis é uma abordagem de interesse ao controle de qualidade daquela matriz complexa visando à sua discriminação em função dos efeitos da sazonalidade sobre sua composição química, com importantes desdobramentos quanto às suas aplicações terapêuticas e industriais.

1.3.1 Controle de qualidade de extratos hidroalcoólicos de própolis – classificação de amostras de própolis em função dos efeitos de sazonalidade

Própolis é o termo genérico aplicado a uma complexa mistura de substâncias, elaborada por abelhas *Apis mellifera* L. a partir de produtos de seu

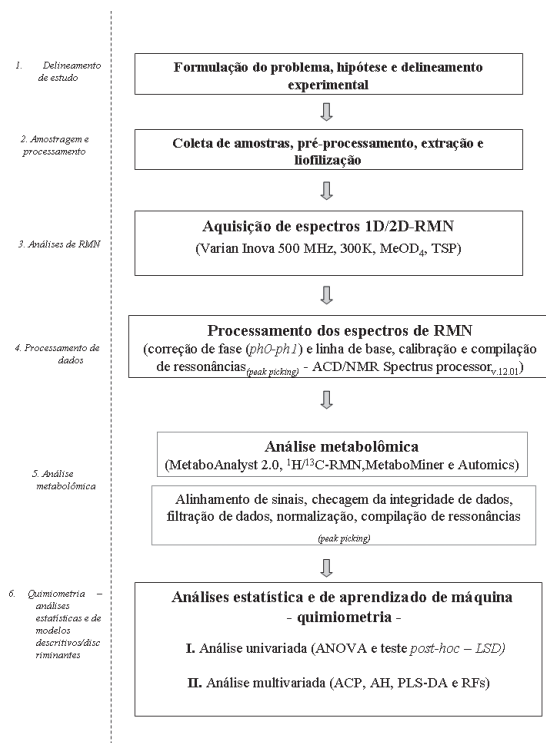


Figura 1.1 Fluxograma das etapas do modelo de estudo para a análise metabolômica da própolis produzida em Santa Catarina, 2010.

próprio metabolismo, isto é, as ceras, de material resinoso e balsâmico coletado de ramos, flores, brotos e exsudados da flora local e também de biomassas (pólen, por exemplo) introduzidas durante sua produção. Estima-se que em torno de 4% a 15% de compostos voláteis e das substâncias coletadas na flora sofram algum tipo de modificação na sua estrutura química por ação do aparato enzimático salivar das abelhas^{42,43}. A própolis possui uma composição química bastante complexa, sendo os constituintes solúveis os materiais cerosos (em média 30%), bálsamos, óleos essenciais e derivados fenólicos. Mais de 300 compostos de baixo peso molecular já foram identificados naquela biomassa, dentre eles: flavonoides, ácidos aromáticos, ácidos graxos, fenóis, aminoácidos, vitaminas (A, B1, B2, B6, C e PP) e minerais (Mn, Cu, Ca, Al, Si, V, Ni, Zn e Cr)⁴⁴.

O uso da própolis é documentado desde as mais remotas civilizações até os tempos atuais. No Egito Antigo, por exemplo, foi utilizada para embalsamar os mortos, e durante a Segunda Guerra Mundial foi adotada em clínicas

soviéticas para a cicatrização de lesões⁴⁵, bem como para o tratamento da tuberculose, observando-se a regressão dos problemas pulmonares e recuperação do apetite⁴⁶. Relatos empíricos sobre o uso de extratos de própolis na medicina popular associados a estudos científicos demonstram ser a própolis um promissor agente terapêutico, devido às suas propriedades antibacterianas, antifúngicas, antivirais, anti-inflamatórias, hepatoprotetoras, antioxidantes, antitumorais e imunomoduladoras⁴⁷. Atualmente, a própolis é um importante produto na medicina alternativa e complementar dos países europeus, Américas do Sul e do Norte e principalmente do Japão⁴⁸.

As primeiras análises químicas de própolis foram realizadas com amostras originárias da Europa e demonstraram a presença de flavonoides agliconas, ácidos fenólicos e seus ésteres. Em 1926, Jaubert constatou a origem vegetal da própolis ao isolar crisina dessa matriz complexa, sendo esse flavonoide também encontrado na espécie *Populus nigra* L. var. *pyramidalis*, a qual vegetava no local da colmeia⁴⁹. No início dos anos 1990, a própolis dos trópicos atraiu a atenção dos cientistas, os quais concluíram se tratar de um produto de composição química bastante distinta comparativamente às amostras europeias. Estes achados indicavam que comparar própolis de regiões produtoras geograficamente discrepantes poderia ser o mesmo que comparar extratos de plantas de distintas famílias⁵⁰. Assim, ainda que algumas atividades biológicas da própolis lhe sejam intrínsecas, isto é, as atividades antioxidantes e antibacterianas, sua composição química é altamente variável em função da flora apícola da região produtora, do componente genético da abelha, da sazonalidade e do método de coleta utilizado, por exemplo. Tais variáveis determinam em alguma extensão as eventuais discrepâncias de atividades farmacológicas de amostras daquela biomassa⁵¹, com efeitos óbvios sobre a qualidade do insumo das indústrias farmacêutica e de cosméticos, por exemplo. Uma consequência resultante desse cenário refere-se à compulsoriedade da caracterização química da própolis em estudos de sua atividade biológica e no desenvolvimento de produtos de interesse à saúde humana e animal. Além disso, para efeitos de aceitação oficial dos sistemas de saúde, em alguns países a comercialização de produtos derivados da própolis requer a padronização química dos extratos daquela matriz como forma de garantir a qualidade, eficiência e segurança do produto final⁵⁰.

Devido à grande biodiversidade dos trópicos, a identificação da(s) espécie(s) vegetal(is) fonte(s) de resina(s) da própolis tende a ser mais complicada em relação às zonas temperadas⁵². Em um ambiente tão biodiverso quanto o território brasileiro, as possibilidades de ofertas de resinas às abelhas são

inúmeras, gerando insumos com ampla gama qualitativa e quantitativa de efeitos biológicos.

A literatura demonstra a tipificação de amostras de própolis no Brasil oriundas de regiões geograficamente distintas, associando-as a atividades biológicas específicas, via uso de marcadores bioquímicos. A própolis brasileira foi classificada em doze tipos segundo o perfil químico obtido via técnicas espectrofotométricas e cromatográficas, além da avaliação das atividades antimicrobianas e antioxidantes⁵². Foram detectadas diferenças qualitativas e quantitativas na composição química e atividades biológicas das amostras estudadas, sendo que os tipos 3 (região Sul), 6 (região Nordeste) e 12 (região Sudeste) apresentaram maior atividade inibitória do crescimento de *Streptococcus mutans*^{52,53}. Estudos adicionais revelaram uma composição química distinta das própolis tipo 6 em relação aos demais tipos, principalmente pela completa ausência de flavonoides e a presença de compostos de natureza mais apolar, isto é, ácidos graxos^{52,54}.

Em trabalho similar visando padronizar a própolis brasileira, propôs-se⁵⁵ o uso de alguns ácidos fenólicos como marcadores bioquímicos em sistemas de análise cromatográfica (cromatografia líquida de alta eficiência – CLAE), classificando as própolis das regiões Sul e Sudeste em quatro tipos principais: BRG, BRP (PR), BRP e BRPG. A própolis do tipo BRG apresenta como marcadores os compostos vanilina, 3-metoxi-4-hidroxicinamaldeído e 2-[1-hidroximetil]-vinil-6-acetil-5-hidroxicumarano, sendo encontrada nos estados do Paraná, Santa Catarina e Rio Grande do Sul. O tipo BRP (Paraná) caracteriza-se pela presença de ácido 3-prenil-4-hidroxicinâmico (PHCA), 9-*E*- e 9-*Z*-2, 2-dimetil-6-carboxietenil-8-prenil-2H-1-benzopirano (DCBEN) e ácido 2,2-dimetil-8-prenil-2H-1-benzopirano-6-propenóico (DPB), sendo típica desse estado. A própolis BRP, originária de São Paulo e Minas Gerais, apresenta como marcadores principais os ácidos cafeico (CAF), 3,5-diprenil-4-hidroxicinâmico (DHCA) e o ácido *p*-cumárico, enquanto a própolis BRPG constituiu uma interface entre os tipos BRP (Paraná) e BRG, caracterizando-se pela presença de compostos desses dois grupos. Embora haja uma grande variabilidade de composição química entre as amostras estudadas, constatou-se que a própolis brasileira apresenta um marcador principal, o DHCA, também denominado de Artepillin C®, um composto de reconhecido efeito antitumoral patenteado pelos japoneses.

A padronização de extratos de própolis a partir do uso de marcador(es) bioquímico(s) associados à origem geográfica daquela biomassa deverá considerar que este(s) é(são) obrigatoriamente constituinte(s) químico(s) da flora local da região produtora, um aspecto que não raro constitui-se em

impedimento técnico à adoção de um marcador bioquímico de abrangência universal⁵⁰. Assim, se o desenvolvimento de produtos eficazes, seguros e de qualidade derivados da própolis tem como pressuposto compulsório o estabelecimento de perfis de composição química determinantes da(s) atividade(s) farmacológica(s) específica(s), a padronização dos extratos de própolis numa dimensão geográfica mais restrita, isto é, regional, se mostra mais credível comparativamente ao âmbito estadual ou nacional, assumindo-se os efeitos dos fatores (a)bióticos típicos sobre o conteúdo de compostos bioativos-alvos, por exemplo flavonoides⁵⁶, derivados do ácido cinâmico e seus ésteres^{42,57} e diterpenos^{57,58}. De fato, em um país de dimensões continentais e mega-biodiverso como o Brasil, estratégias de padronização química/atividade(s) biológica(s) dos extratos da própolis em âmbito nacional não se mostram adequadas como suporte ao setor produtivo, conforme previamente observado^{44,46,55}. Por fim, ressalta-se que para além dos trabalhos de caracterização química da própolis associada às atividades biológicas de interesse, estudos demonstram a viabilidade da aplicação de abordagens analíticas químicas, usuais em estudos metabolômicos, focando a determinação da origem geográfica daquela biomassa e subsidiando o requerimento de registro de indicação geográfica desta^{52,59}, ainda que em número pouco expressivo, considerando a importância terapêutica/profilática e econômica da própolis^{60,61}.

A própolis proveniente da Europa é a mais estudada e conhecida, tanto do ponto de vista químico quanto farmacológico, caracterizando-se por uma alta concentração de flavonoides⁵⁰. Um dos compostos fenólicos mais estudados naquela biomassa é o éster fenílico do ácido cafeico, que foi identificado como seu principal composto antifúngico e bactericida⁶², apresentando também outras propriedades farmacológicas importantes como a antitumoral⁶³, a antioxidante⁶⁴ e a propriedade de inibir a síntese da integrase VIH-1, envolvida na replicação do vírus da Síndrome da Imunodeficiência Adquirida (AIDS)⁶⁵. Os diversos achados em relação à própolis europeia levaram ao desenvolvimento de um método de tipificação dessa matriz, por meio da análise de sete compostos fenólicos (isto é, *target metabolites*) por cromatografia líquida de alta eficiência (CLAE), seguida da determinação espectrofotométrica UV-Vis (espectrofotometria ultravioleta-visível) dos conteúdos totais de flavonas e flavonóis, flavonovas e dihidroflavonóis e compostos fenólicos. Essa abordagem discrimina se dada amostra de própolis é proveniente dos exsudatos dos meristemas de álamo^{66,67}.

Embora os flavonoides sejam os principais componentes da própolis europeia, o mesmo não se pode dizer da própolis brasileira, na qual

quantidades inferiores daqueles metabólitos secundários parecem ocorrer⁵⁵. De fato, em geral a própolis brasileira é caracterizada principalmente pela presença de poucos flavonoides, por exemplo, canferide 5,6,7-triidroxi-3,4'-diidroxiflavona, aromadendrina-4-metil éter⁶⁸, pinobanksina⁶⁹, crisina e galangina⁵⁵. Em amostras provenientes da região Nordeste, por exemplo, detectou-se a completa ausência de flavonoides e a presença de compostos de natureza mais apolar, como os ácidos graxos⁵². Ressalta-se que esse tipo de própolis mostrou-se mais efetiva na inibição do crescimento de *Streptococcus mutans* e *Staphylococcus aureus*⁷⁰, o que leva a supor que a ausência de flavonoides não significa necessariamente uma própolis de qualidade inferior. De outra forma, parece haver um predomínio de compostos derivados de ácidos fenólicos com atividades biológicas importantes em extratos de própolis brasileira, a saber: ácido 3-prenil-4-hidroxicinâmico (PHCA), ácido 2,2-dimetil-8-prenil-2H-1-benzopirano-6-propenóico (DPB), 9-E- e 9-Z-2, 2-dimetil-6-carboxietenil-8-prenil-2H-1-benzopirano (DCBEN) e 3,5-diprenil-4-hidroxicinâmico (DHCA)^{71,72}.

Por essa razão, dependendo do objetivo do estudo, a caracterização da própolis tropical de maneira similar à europeia pode conduzir a erros, uma vez que as técnicas de CLAE e a espectrofotometria UV-vis não são adequadas à análise de ceras e terpenoides⁷³, por exemplo. Numa abordagem analítica alternativa à CLAE-UV-Vis, amostras de própolis cubanas foram investigadas em sua composição química via ¹H- e ¹³C-RMN, CLAE e CLAE-EM. Os resultados permitiram a classificação daquela biomassa em três grupos consoante aos seus marcadores bioquímicos principais. A própolis cubana marrom caracterizou-se pela presença de benzofenona poli-isopreniladas (em especial nemorosona) e seus derivados, enquanto isoflavonoides, associados ou não a componentes alifáticos (terpenoides e esteróis) foram típicos da própolis vermelha. Por sua vez, a própolis amarela apresentou perfis espectrais com ressonâncias predominantes na região alifática isto é, 0,00-3,00 ppm⁷⁴. Em seu conjunto, os resultados sugerem que a adoção de uma ferramenta analítica não seletiva, isto é, RMN, para a obtenção de perfis metabólicos mostrou-se adequada à discriminação de tipos de própolis, por sua habilidade em detectar simultaneamente compostos fenólicos, ceras e terpenoides. Mais recentemente⁷⁵, em abordagem analítica similar (1D e 2D-RMN) foram identificados perfis metabólicos de amostras de própolis do Sul do Brasil (estado de Santa Catarina) com sinais típicos na região alifática, associados a grupos funcionais acetil (COCH₃, 2,07-2,37 ppm) e da fração alceno (C-CH₂-C, 1,30-1,33 ppm; C-CH-C, 1,21 e 1,47 ppm) de ácidos graxos e ceras comumente encontrados naquela biomassa.

Em função do exposto, no contexto de estudos metabolômicos, observa-se que a análise de matrizes complexas, isto é, própolis, pode focar em compostos-alvo de interesse específico (*target metabolites*), requerendo, portanto, técnicas analíticas seletivas, ou ainda adotar uma abordagem mais ampla, em que se busca identificar o máximo possível de compostos químicos a partir de técnicas não seletivas (*metabolic profile*). Nesse último caso, o conjunto de dados gerados permite a obtenção de informações adicionais associadas a padrões metabólicos, quando aplicado à construção de modelos de descrição e classificação através de técnicas estatísticas multivariadas (*metabolic fingerprint*) – Tabela 1.1. Independentemente da abordagem analítica de investigação metabolômica e considerando-se as características intrínsecas da(s) técnica(s) adotada(s), ressalta-se a necessidade de padronização e estudos de validação intra/interlaboratorial do(s) método(s) utilizado(s), conferindo-lhe(s) a necessária confiabilidade e robustez³².

A partir de um cenário de escassas informações relativas à composição química da própolis originária de Santa Catarina, adotou-se uma estratégia de estudos metabolômicos constituída por duas etapas: i) determinação do perfil metabólico via técnicas analíticas não seletivas (1D/2D-RMN e FTIR) associadas à construção de modelos descritivos e de classificação à geração de padrões de reconhecimento da origem geográfica amostral (*metabolic fingerprint*); e ii) análises de compostos-alvos de elevado valor agregado detectados a partir dos perfis metabólicos, através de técnicas analíticas seletivas (CLAE-UV-Vis, CLAE-EM). A hipótese de trabalho considera que a sazonalidade altera o perfil metabólico da própolis catarinense, em função da ação dos fatores climáticos sobre a flora apícola. Na sequência, serão descritos com algum detalhamento os procedimentos experimentais adotados na etapa (i) das análises metabolômicas, com ênfase nas análises de espectroscopia de RMN.

1.4 PROTOCOLO EXPERIMENTAL – PERFIS METABÓLICOS E PADRÕES DE RECONHECIMENTO DE AMOSTRAS DE PRÓPOLIS CATARINENSE

1.4.1 Material amostral – coleta e preparo

No ano de 2010, amostras de própolis catarinense (100 g/amostra, n = 59, 16 cidades) foram coletadas no outono (AU, abril, n = 15), inverno

(WI, julho, $n = 13$), primavera (SP, outubro, $n = 15$) e verão (SM, janeiro, $n = 16$), em cinco regiões geográficas do estado de Santa Catarina (leste, $n = 7$; central, $n = 7$; serrana, $n = 12$; norte, $n = 7$; e oeste, $n = 26$). Um aparato denominado coletor inteligente foi previamente instalado nas colmeias como estratégia de uniformização dos procedimentos de coleta, dado aos reconhecidos efeitos que esta prática tem sobre o perfil químico da própolis⁷⁶. Imediatamente após a coleta, as amostras foram acondicionadas em embalagens plásticas, seladas, identificadas e transferidas ao laboratório. Um procedimento padrão de limpeza e seleção das amostras foi executado, seguido do armazenamento destas em embalagens plásticas a $-18\text{ }^{\circ}\text{C}$ até posterior análise.

1.4.2 Extração

A obtenção dos extratos de própolis utilizou uma solução de etanol: água (70:30, v/v), por ser aquele solvente orgânico o mais utilizado à produção comercial de extratos de própolis⁷⁶. A um volume de 25 ml de solução extratora, 2 g de própolis foram adicionados, seguido de maceração em gral e pistilo com o auxílio de N_2 líquido, e incubação a $70\text{ }^{\circ}\text{C}/30\text{ min.}$, sob agitação constante e ausência de luz. O extrato hidroalcoólico (EH) foi coletado, centrifugado (5.000 rpm/5 min.), recuperando-se o sobrenadante e corrigindo-se o volume final a 25 ml com a solução extratora, seguido de liofilização.

1.4.3 Espectroscopia de RMN uni e bidimensional

O EH liofilizado (50 mg) foi adicionado de 700 μl de metanol deuterado (MeOD_4) e centrifugado (5.000 rpm/10 min.), coletando-se o sobrenadante (650 μl) e transferindo-o a tubo de RMN com 5 mm de diâmetro interno. Como padrão interno, utilizou-se o sal sódico do ácido 3-(trimetilsilil)-propiónico- d_4 (sal sódico do ácido 3-(trimetilsilil)-propiónico- d_4 deuterado – TSP d_H 0,00 ppm – 0,4 mM, pH = 3,8)⁷⁷.

Para obter espectros comparáveis, estes foram obtidos utilizando-se os mesmos parâmetros de configuração e considerando-se um período de um minuto para a estabilização da amostra na sonda de detecção, previamente à aquisição dos dados.

Os espectros de alta resolução foram adquiridos em equipamento Varian Inova 500 MHz, equipado com sonda BBI (5 mm) de detecção inversa e

bobinas geradoras de gradiente de campo (coordenada z), operando em frequência de ressonância de 500,13/125,03 MHz (11,7 Tesla) de $^1\text{H}/^{13}\text{C}$. Os parâmetros de aquisição dos espectros de ^1H -RMN foram configurados após ensaios preliminares, conforme segue: tempo de aquisição de 4 s (*acquisition time*), tempo de espera 2 s (*delay time*) e tempo de reciclagem 6 s (*recycle time*), ao longo de uma janela espectral de 8000 Hz, com supressão do sinal de água (4,87 ppm, sequência de pulso Watergate, extensão de pulso 8,5 μs – 90°, nível de potencial 50 dB, 300 K) e tempo de mistura 12 ms (*mixing time*).

O tempo de reciclagem de 6 s mostrou-se adequado, evitando significativamente (< 10%) a saturação de sinais, isto é, $3T_1$ s. Para cada amostra, 32/2.000 varreduras ($^1\text{H}/^{13}\text{C}$, respectivamente) foram realizadas, coletando-se 16 K (^1H) e 32 K (^{13}C) pontos (*time domain*) em modo de intensidade absoluta. Os dados espectrais foram processados utilizando-se 64 K pontos, através de uma rotina implementada no programa ACD/Spectrus processor (v.12.01 – Advanced Chemistry Development, Toronto, Ontario, Canada), consistindo de multiplicação exponencial ($\text{lb} = 0,5$ Hz), ajuste de fase automático (*ph0* e *ph1*), correção da linha de base e calibração manual em relação ao sinal do padrão interno (TSP d_H 0,00 ppm).

Os experimentos de 2D-RMN incluíram registros de correlação total ($^1\text{H}/^1\text{H}$, TOCSY), de correlação heteronuclear ($^1\text{H}/^{13}\text{C}$, HSQC) e eventualmente de difusão ordenada (DOSY) e de correlação por acoplamento J (“ J -resolved Spectroscopy”), sendo que somente os parâmetros das duas primeiras abordagens experimentais de 2D-RMN serão descritos. Os espectros de TOCSY foram obtidos em modo de fase sensível, com incremento de fase proporcional ao tempo (TPPI) e sequência de pulso *wgtocsy* (largura espectral de 7992,19 Hz em F1 e F2, 1024-F1/512-F2 pontos de dados e tempos de aquisição, associação e de relaxação de 0,1280 s-F1/0,0320 s-F2, 100 ms e 1,5 s, respectivamente).

Os espectros de correlação heteronuclear $^1\text{H}/^{13}\text{C}$ (HSQC) foram adquiridos utilizando-se a sequência de pulso *gChsqc* e os parâmetros que seguem: número de varreduras 8, varreduras fictícias 4 (*dummy scans*), número de pontos (*time domain*), 1.024 [F2 – ^1H] e 512 [F1 – ^{13}C], janela espectral de 7.992,19 Hz (F2) e 21.080,12 Hz F1, resolução digital 1,8777 Hz (F2) e 200,020 Hz (F1), tempo de aquisição 0,1280 s-F2/0,0121-F1, tempo de espera 0,6 s. A escolha dessa técnica de correlação heteronuclear com o desacoplamento dos sinais de ^{13}C baseia-se no fato de que ela propicia um número elevado de sinais bem definidos e não requer o ajuste de fase no processamento dos dados, evitando a inserção de dados espúrios durante o

cálculo e processamento espectral⁷⁶. As configurações utilizadas permitiram a obtenção de espectros bidimensionais com boa resolução, reduzindo significativamente as dificuldades de associação das ressonâncias aos metabólitos derivadas da sobreposição de sinais tipicamente observada em espectros de ¹H-RMN⁷⁸.

Os dados espectrais de 2D-RMN foram processados utilizando-se uma rotina implementada no programa ACD/Spectrus processor (v.12.01 – Advanced Chemistry Development, Toronto, Ontario, Canada), considerando a calibração dos espectros a partir do sinal do padrão interno (TSP, d_{1H e 13C} 0,00 ppm) e a correção da linha de base. Adicionalmente, aos dados espectrais de HSQC em F1 e F2 foram aplicadas durante o processamento uma função e uma constante quadrática sinoidal⁴⁰.

1.4.4 Análise de dados espectrais de 1D e 2D-RMN

O uso da RMN de alta resolução uni e bidimensional nas amostras de própolis gera uma grande quantidade de informações, as quais podem ser extraídas do conjunto de dados recolhido. No presente estudo, estes correspondem a 59 amostras, tendo sido coletadas três réplicas técnicas por amostra na fase de aquisição dos espectros de RMN. Dado que as réplicas se mostraram muito fiéis entre si, com um grau de variabilidade muito reduzido, usou-se na análise de dados o espectro médio para cada amostra, sem perda de informação.

A análise dos dados espectrais de 1D-RMN utilizou o *software* MetaboAnalyst^{79*}. Esse *software* tem-se afirmado nos últimos anos como uma boa opção ao tratamento de dados de metabolômica, executando uma ampla gama de operações de pré-processamento de dados e de análises uni e multivariadas. Além disso, é uma ferramenta de fácil utilização e acesso, dada a sua disponibilidade a partir de um *web browser* comum.

Em primeiro lugar, a matriz de dados espectrais de ¹H-RMN foi pré-processada, isto é, os espectros coletados (0,20 ppm a 13,00 ppm) foram convertidos em uma lista de picos cuja primeira coluna refere-se ao deslocamento químico (ppm) e a segunda à intensidade do sinal. No total, foram obtidos 25.403 picos, representando uma média de cerca de 430 picos por amostra. Com base nessas informações, procedeu-se ao alinhamento dos picos entre as amostras, de forma a identificar os picos comuns. Nessa operação, foi

* Disponível livremente em versão web em <http://www.metaboanalyst.ca/MetaboAnalyst>.

usado o algoritmo implementado no MetaboAnalyst por omissão, tendo-se usado uma janela móvel de largura 0,03 ppm e um passo de 0,015 ppm. Os picos do mesmo grupo foram alinhados a sua mediana considerando todas as amostras, sendo que ressonâncias com frequência de ocorrência inferior a 25% em cada grupo foram excluídas. Como resultado, foram obtidos 239 picos, em uma matriz de 59 × 239 pontos de dados para posterior análise. O passo adicional constou do tratamento dos valores omissos, tendo sido estes substituídos por um valor de 0,00005, isto é, metade do valor positivo mais baixo permitido nos dados originais, sendo este valor assumido como limite da detecção. Subsequentemente, os dados foram normalizados através da transformação logarítmica destes.

Na etapa seguinte, os dados pré-processados foram submetidos à análise univariada, tendo em conta a sua divisão nas estações de coleta (AU, WI, SP e SM). Esses métodos são comumente utilizados à exploração preliminar dos dados, fornecendo uma visão dos atributos (isto é, ressonâncias) mais relevantes na discriminação das condições em estudo.

Dada à existência de mais de dois tratamentos (estações), realizou-se a análise de variância (ANOVA), seguida de testes estatísticos *post hoc* (teste LSD de Fisher), permitindo identificar os sinais como maior variabilidade nas suas intensidades entre os tratamentos (estações), bem como os pares de tratamentos nos quais essa variabilidade se mostrou significativa, apresentando o respectivo *p*-value. Os picos mais importantes identificados são mostrados na Tabela 1.2, na qual claramente se nota o predomínio e a significância de compostos com estrutura anomérica (3,00 ppm a 5,50 ppm) nos agrupamentos detectados.

Tabela 1.2 Ressonâncias (1H-RMN, ppm) identificadas via análise de variância (ANOVA). A primeira, a segunda e a terceira colunas representam a ressonância selecionada, o logaritmo dos valores de *p* (*p*-value) e os pares de tratamentos para os quais a diferença é significativa (*p* < 0,05 – teste LSD), respectivamente

RESSONÂNCIA (PPM)	LOG ₁₀ (P-VALUE)	TESTE LSD FISHER
4,68	24,393	SM* - AU; SM - SP; SM - WI
4,65	19,398	SM - AU; SM - SP; SM - WI
4,58	16,470	SM - AU; AU - SP; AU - WI; SM - SP; SM - WI
4,32	14,087	SM - AU; AU - SP; AU - WI; SM - SP; SM - WI
4,54	13,731	SM - AU; AU - SP; AU - WI; SM - SP; SM - WI

RESSONÂNCIA (PPM)	LOG ₁₀ (P-VALUE)	TESTE LSD FISHER
4,06	13,338	AU - SP; AU - WI; SM - SP; SM - WI
4,71	12,336	SM - AU; SM - SP; SM - WI
3,45	11,844	AU - SM; AU - SP; AU - WI
3,39	11,677	AU - SM; AU - SP; AU - WI
4,18	11,572	AU - SP; AU - WI; SM - SP; SM - WI
3,42	11,395	AU - SM; AU - SP; AU - WI
4,15	11,072	AU - SP; AU - WI; SM - SP; SM - WI
4,62	10,737	SM - AU; SM - SP; SM - WI
4,5	10,708	AU - SP; AU - WI; SM - SP; SM - WI
4,26	10,359	AU - SP; AU - WI; SM - SP; SM - WI
4,47	10,296	SM - AU; AU - SP; AU - WI; SM - SP; SM - WI
4,29	10,270	AU - SP; AU - WI; SM - SP; SM - WI
4,38	9,797	AU - SP; AU - WI; SM - SP; SM - WI
4,44	9,793	AU - SP; AU - WI; SM - SP; SM - WI
4,03	9,259	AU - SP; AU - WI; SM - SP; SM - WI

* Legenda: SM, verão; AU, outono; SP, primavera, WI, inverno.

Em casos nos quais a estrutura dos dados permite a divisão das amostras em apenas dois grupos (tratamentos), é possível realizar uma análise semelhante à anterior com testes estatísticos baseados na distribuição *t*-student, também disponibilizados pelo *software* MetaboAnalyst.

Uma perspectiva mais completa sobre os dados requer a identificação dos metabólitos presentes em cada amostra, com base em seus sinais de ressonâncias (δ , ppm). Essa operação, por sua vez, requer a análise visual dos espectros, ainda que tal procedimento exija certo grau de conhecimento do usuário e disponibilidade de tempo. O resultado desse processo para os dados referidos anteriormente é mostrado na Tabela 1.3, no qual dezenove metabólitos foram identificados nos EHs, ainda que de forma não homogênea no material amostral, sugerindo o efeito de sazonalidade. A atribuição das ressonâncias aos metabólitos usualmente emprega informações disponíveis em bases de dados (algumas são referidas na Tabela 1.4), literatura

especializada e ferramentas de bioinformática, como, por exemplo, as plataformas MetaboMiner* ou Automics**. Dessa forma, foi possível identificar metabólitos de natureza de ácidos orgânicos, aminoácidos, álcoois, monossacarídeos e compostos aromáticos, por exemplo, eventuais candidatos a marcadores bioquímicos da origem da própolis, consoante às regiões geográficas em estudo.

A título de exemplo, dois espectros de ¹H-RMN de amostras de própolis outonais (2010), originárias das regiões serrana (município de São Joaquim) e oeste (município de Descanso) de Santa Catarina, são mostrados nas figuras 1.2a e 1.2b, com destaque para as janelas espectrais de ocorrência de compostos de estruturas alifática (0,00 ppm a 3,00 ppm), anomérica (3,00 ppm a 5,50 ppm) e aromática (5,50 ppm a 8,50 ppm).

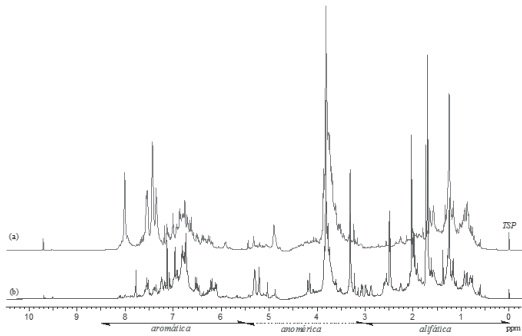


Figura 1.2 Perfis espectrais de ¹H-RMN (500 MHz, MeOD₄, TSP dH 0,00 ppm, pH = 3,8) de amostras de EH de própolis catarinense originárias das regiões (a) serrana (São Joaquim) e (b) oeste (Descanso), coletadas no outono de 2010.

Tabela 1.3 Deslocamentos químicos de ¹H e ¹³C, multiplicidade de prótons e assinalamentos (500 MHz) de metabólitos de baixo peso molecular (< 1.500 Da) identificados em amostras de própolis catarinense (TSP dH 0,00 ppm – 0,4 mM, pH = 3,8)

METABÓLITOS	δ, ppm ¹ H (MULTIPLICIDADE, ASSINALAMENTO) – δ, ppm ¹³ C *
Ácido b-OH/butírico	1,21 (d); 2,28 (m); 2,52 (m) – 47,44
Alanina	1,47 (d, bCH ₃) – 17,52; 4,10 (aCH) – 51,10
Arginina	2,02 (m, bCH ₂) – 31,72; 3,32 (dCH ₂) – 42,29; 4,10 (t, aCH) – 54,87
Ascorbato	3,73 (m) – 62,06; 4,01 (m) – 69,06; 4,50 (d) – 77,90

* Disponível em <<http://wishart.biology.ualberta.ca/metabominer/>>.

** Disponível em <<https://code.google.com/p/automics/>>.

METABÓLITOS	δ , ppm ^1H (MULTIPLICIDADE, ASSINALAMENTO) – δ , ppm ^{13}C *
Formiato	8,43 (s, HCOOH)
Fumarato	6,52 (s) – 136,08
Ácido gálico	7,07 (s, C2H, C6H-ring) – 117,36
Citrato	2,84 (dd, a, g CH) – 44,28
b-Fructose	3,58 (C1H) – 62,65; 3,82 (C5H) – 81,00; 4,10 (C3H) – 75,40
α -Glucose	3,51 (C2H) – 71,90; 3,84 (C5H) – 71,17; 5,22 (d, C1H) – 94,20
b-Glucose	3,23 (C2H) – 76,25; 3,41 (C4H) – 71,37; 4,62 (d, C1H) – 98,10
Glicina	3,85 (s, αCH_2) – 44,29
Isoleucina	0,89(t) – 11,54; 0,96 (t) – 14,02; 3,77 (m) – 61,55
Lisina	1,49 (m) – 21,10; 1,72 (m) – 26,01; 1,92 (m) – 31,66; 3,06 (t) – 40,01
Melibiose	3,27 – 75,20; 3,56 – 73,76; 3,72 – 73,87; 3,83 – 69,78; 3,91 – 70,97; 4,00 – 71,97; 5,00 – 98,92
Ornitina	1,81 (m) – 25,00; 2,01 (m) – 29,55; 3,06 (t) – 42,18
Prolina	2,06 (m) – 25,98; 2,18 (m) – 30,10
Piruvato	2,36 (s) – 29,01
Quercetina	6,18 (d) – 98,04; 6,87 (d) – 116,11; 7,55 (d) – 119,94; 9,12 (s) – 135,97
Succinato	2,39 (s) – 33,94
Sacarose	3,53 (G2H, dd) – 71,07; 3,67 (s, F1H) – 62,66; 5,42 (G1H, d) – 93,25
Tartarato	4,33 (s) – 72,68
Triptofano	3,23 (dd) – 28,30; 7,19 (d) – 122,40; 7,33 (s) – 128,20; 7,54 (d) – 114,80
5-uridina monofosfato	4,10 – 67,00; 4,22 – 76,66; 4,31 – 71,90; 6,94 – 118,22; 5,90 – 105,21
α -Xilose	3,52 (dd) – 73,77; 3,63 – 72,91; 3,68 – 61,92; 5,19 (d) – 93,25

* Assinalamentos determinados via ^1H -RMM, TOCSY e HSQC.

Fonte: Maraschin *et al.* (2012) e dados não publicados pelos autores.

Tabela 1.4 Bases de dados de referência em estudos metabolômicos

BASE DE DADOS	URL OU ENDEREÇO NA INTERNET	OBSERVAÇÕES
Database of biochemical, genetic, and genomic metabolic network reconstruction (BiGG)	http://bigg.ucsd.edu/	Repositório de dados de metabólitos de bactérias, leveduras e humanos e vias metabólicas – ênfase em biologia sistêmica
Biological Magnetic Resonance Data Bank (BioMagResBank)	http://www.bmrb.wisc.edu/metabolomics/	Repositório de dados de RMN de macromoléculas (proteínas, peptídeos, ácidos nucleicos) e metabólitos
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/kegg1.html	Base que integra dados genômicos, químicos e de biologia funcional

BASE DE DADOS	URL OU ENDEREÇO NA INTERNET	OBSERVAÇÕES
Fiehn metabolome database	http://fiehnlab.ucdavis.edu/db	Listagem de metabólitos com identificadores (IDs), imagens, sinônimos e conexão com a base de dados KEGG
Golm metabolome – Comprehensive Systems Biology Database	http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html	Plataforma de dados de EM e abordagens experimentais em perfis metabólicos e bioinformática
Human metabolome database	http://www.hmdb.ca/	A maior e mais completa base de dados metabolômicos dedicada à espécie humana
NIST spectral database	http://webbook.nist.gov/chemistry/	Base de dados espectrais (EM, FTIR, NIR RMN) e de propriedades químicas e físicas de moléculas
Plant Metabolic Network (PMN)	http://plantcyc.org/	Banco de dados de genes, enzimas, metabólitos e vias metabólicas primárias e secundárias de vegetais
Platform for RIKEN Metabolomics (PRiMe)	http://prime.psc.riken.jp/	Banco de dados espectrais (EM, EM/EM, RMN) e de ferramentas de integração de informações metaboloma/transcriptoma
Scripps Center for Metabolomics – METLIN: Metabolite and Tandem MS Database	http://metlin.scripps.edu/	Repositório de dados de EM em tandem (EM/EM) e metabólitos
Spectral Database for Organic Compounds (SDBS)	http://sdb.sdb.aist.go.jp/	Banco de dados integrado de espectros (EM, FTIR, Raman, RMN e ressonância de spin eletrônico – RSE)

1.4.5 Quimiometria

Os dados de metabolômica, como já foi demonstrado na Seção anterior, têm características multivariadas, isto é, cada amostra é caracterizada por diversos valores, nesse caso por um grande número de ressonâncias. Assim, naturalmente, os métodos mais eficientes para a exploração e extração das informações desses dados são métodos de análise multivariada.

Nesse contexto, o primeiro método aplicado aos dados espectrais de 1D-RMN foi a Análise de Componentes Principais (ACP). Trata-se de um método não supervisionado, isto é, que não requer a definição *a priori* de grupos para as amostras, sendo usado essencialmente como um método de redução de dimensionalidade. A ACP, criada em 1901 por Karl Pearson⁸⁰, consiste em um procedimento algébrico que converte as variáveis originais (que são tipicamente correlacionadas) num conjunto de variáveis não correlacionadas (linearmente), que se designam por componentes principais (CPs) ou variáveis latentes. Assim, a ACP fornece um mapeamento de um

espaço com N dimensões (em que N é o número de variáveis originais) para um espaço com M dimensões (em que M é tipicamente muito menor do que N). As CPs são ordenadas pela quantidade decrescente de variabilidade (variância) que explicam. Cada CP é gerada de forma a explicar o máximo de variabilidade da parte ainda não explicada, devendo ser ortogonal às CPs anteriores. É importante notar que a ACP é sensível à escala dos dados, característica devido à qual se recomenda sua normalização prévia. A ACP consiste numa decomposição dos dados originais (uma matriz X) em duas matrizes: $T.P^T$. A matriz T denomina-se *scores* (distribuição fatorial), indicando as coordenadas dos exemplos iniciais (linhas de X) no novo sistema de coordenadas dado pelas CPs. As CPs determinadas são combinações lineares das variáveis originais, sendo os coeficientes destas no espaço original dados pelas colunas da matriz P e designados por *loadings* (contribuição fatorial). Se forem consideradas apenas as primeiras k componentes principais, isto implicará em considerar também apenas as primeiras k colunas das matrizes T e P , obtendo-se uma aproximação dos dados originais que será tanto mais precisa quanto maior for o valor de k .

A aplicação da ACP é realizada usando o *software* MetaboAnalyst, que por sua vez usa uma função implementada (*prcomp*) sobre o sistema R^{81*}. A Figura 1.3 mostra os *scores* das duas primeiras componentes, bem como

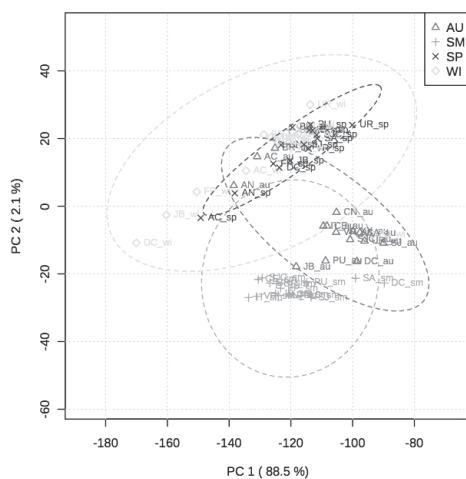


Figura 1.3 Contribuições fatoriais (*scores*) dos componentes principais CP1 e CP2 calculados a partir dos dados espectrais de ¹H-RMN de amostras de própolis catarinense, com o auxílio do *software* MetaboAnalyst.

* Ver <<http://www.r-project.org/>>.

a variância explicada. Note-se que a primeira CP explica mais de 88% da variância amostral. As cores dos pontos revelam os grupos associados às estações do ano, revelando a qualidade da separação gerada por essas duas CPs (notar que essa informação não é usada pelo método).

Uma classe de métodos alternativos, ainda de natureza não supervisionada, está direcionada à tarefa de agrupar as amostras de acordo com suas similaridades. As abordagens para este problema variam sobretudo na forma como se define a similaridade, nos algoritmos usados para realizar o agrupamento e na forma de visualizar e interpretar os resultados. Neste texto, abordar-se-á apenas o *clustering* (agrupamento) hierárquico para os dados em estudo, que constitui a abordagem mais popular para essa tarefa.

O *clustering* hierárquico⁸² tem como principal característica o tipo de resultados que gera. Nesse caso, o resultado final do processo é uma árvore binária que representa possíveis divisões dos dados em *clusters* (grupos). Assim, na raiz, todos os dados estão agrupados num único *cluster*, e ao descer na árvore os *clusters* vão se dividindo de forma binária, ou seja, em cada nó da árvore são criados dois *clusters* pela divisão de um único. O método mais usado para construir essas árvores é designado de aglomerativo, pois se inicia com um *cluster* para cada exemplo (folhas da árvore). Em cada interação, vão se juntando dois *clusters* e criando um único (gerando os nós na árvore), até se atingir o ponto em que todos os *clusters* estão unidos num único. O processo é baseado numa matriz de distâncias em que estão compilados os valores entre todos os pares de objetos. Essa matriz é construída aplicando uma métrica de similaridade sobre as linhas da matriz inicial de

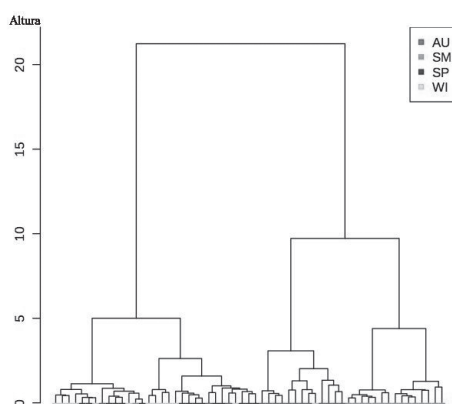


Figura 1.4 *Clustering* hierárquico dos dados espectrais de ¹H-RMN de amostras de própolis catarinense calculado com o auxílio do software MetaboAnalyst. As cores das folhas da árvore denotam os grupos (tratamentos) das amostras.

dados, como, por exemplo, as métricas clássicas de distância euclidiana ou de Manhattan.

O MetaboAnalyst realiza a análise de agrupamentos hierárquicos usando a função (*hclust*) do sistema R e calculando a correlação de Pearson como métrica para a determinação da distância. Os resultados são sumarizados na Figura 1.4, podendo se constatar uma boa separação dos grupos (ainda que essa informação não seja usada pelo método).

Ressalta-se ainda que o processo de *clustering* pode considerar como elemento de agregação os metabólitos (ou no caso em estudo, as ressonâncias). Uma forma de visualizar os dois processos de aglomeração de forma simultânea são os *heatmaps*, conforme exemplificado na Figura 1.5.

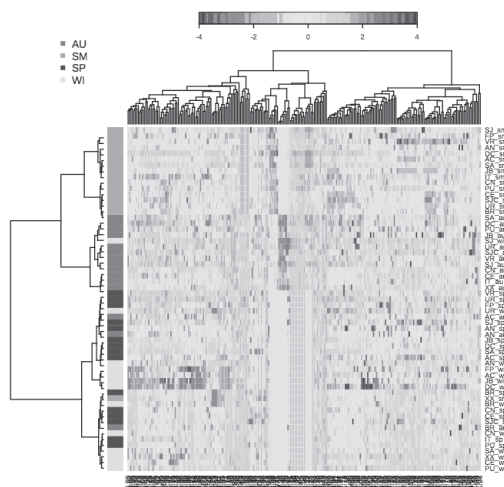


Figura 1.5 *Heatmap* para os dados espectrais de ^1H -RMN das amostras de própolis em estudo. As colunas representam ressonâncias (ppm) e as linhas representam as amostras, sendo as suas cores (lado esquerdo) determinadas pelas estações de coleta (legenda no canto superior esquerdo).

A análise realizada até o momento focou no uso de métodos de natureza não supervisionada, ou seja, pretende-se estudar quais as ressonâncias que distinguem as amostras de grupos (tratamentos) diversos, isto é, estações, mas tal informação não é utilizada na alimentação dos métodos. Adicionalmente, há métodos alternativos a essa abordagem, ditos supervisionados, que permitem usar tal informação para criar melhores modelos discriminativos.

Um dos métodos supervisionados mais utilizados na área da metabolômica é a análise discriminante associada à análise dos quadrados mínimos

parciais (*Partial Least Squares Discriminant Analysis* – PLS-DA). Essa análise representa uma extensão da ACP que tira partido da informação sobre os grupos de cada amostra, procurando maximizar a separação entre grupos de observações⁸³. O método funciona pela descoberta de variáveis latentes nos dados, tal como na ACP, porém busca separar os grupos amostrais não maximizando a variância.

A aplicação do método PLS-DA aos dados espectrais (¹H-RMN) da própolis é ilustrada na Figura 1.6, em que uma boa separação dos grupos de amostras usando apenas três componentes principais pode ser observada. Adicionalmente, a Figura 1.7 mostra uma medida da importância das variáveis (ressonâncias) no modelo de classificação construído, identificando aquelas de maior efeito discriminante, quer de forma genérica, quer para cada grupo (tratamento) em particular. Dado que o PLS-DA é um método que pode incorrer com facilidade em sobreajustamento, uma avaliação rigorosa deve ser realizada. Neste caso, o uso de validação cruzada⁸⁴ demonstrou erros de previsão na ordem dos 10%.

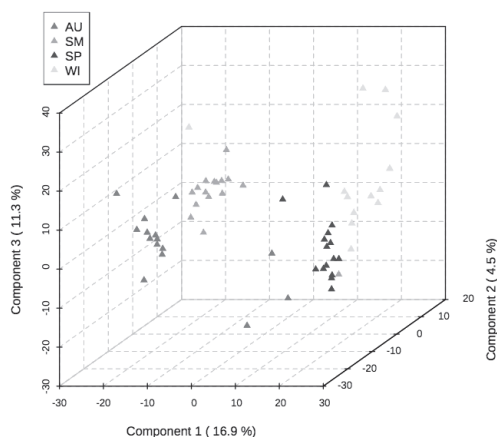


Figura 1.6 Discriminação amostral da própolis catarinense ao longo dos eixos CP1 (16,9%), CP2 (4,5%) e CP3 (11,3%), consoante aos tratamentos (estações) em estudo, a partir da análise discriminante associada ao cálculo dos quadrados mínimos parciais (PLS-DA).

Uma das limitações do método de PLS-DA refere-se à sua dependência da escala dos dados na escolha dos atributos relevantes⁸⁵. Há, porém, métodos alternativos nas áreas da aprendizagem máquina e da mineração de dados em que essa dependência inexistente⁸⁶. Neste estudo, adotou-se a aplicação de árvores de decisão que apresentam em seus nós os atributos (isto é,

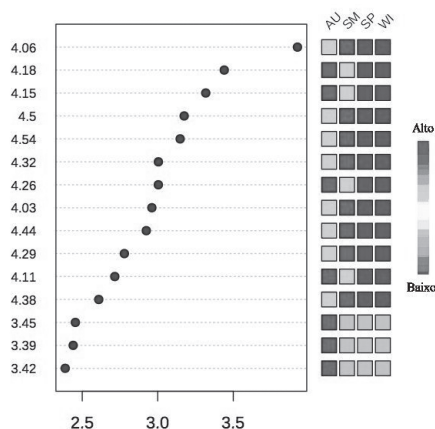


Figura 1.7 Importância das variáveis (ressonâncias, ppm) determinada pelo método PLS-DA para os dados espectrais de ^1H -RMN da própolis catarinense, safra 2010.

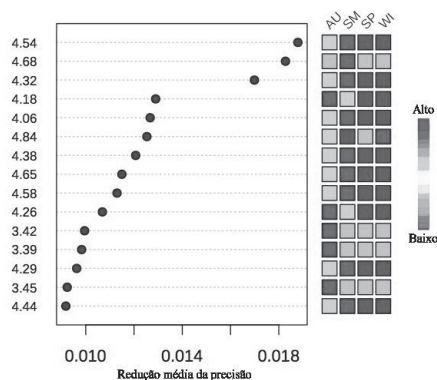


Figura 1.8 Importância das variáveis (ressonâncias) calculadas pelo método RF para os dados espectrais (^1H -RMN) da própolis catarinense, safra 2010.

ressonâncias), enquanto nas folhas tem-se os valores das classes (ou grupos). Mais recentemente, um método que usa conjuntos de árvores de decisão, denominado *random forests* (RFs), vem sendo aplicado à análise metabolômica⁸⁷. Esse método realiza a inferência de centenas de árvores de decisão para um conjunto de dados, sendo posteriormente usadas para prever novos casos via processo de votação baseado nos resultados individuais de cada árvore.

As RFs são também implementadas pelo MetaboAnalyst usando um pacote do sistema R. Aplicando o método aos dados espectrais de 1D-RMN das própolis catarinenses, a estimativa de erro situou-se em torno de 15%. Adicionalmente, o modelo identifica as ressonâncias mais usadas nas árvores de decisão, com boa capacidade preditiva (Figura 1.8).

O conjunto de resultados derivados das análises de PLS-DA e RFs claramente indica ser a região anomérica a de maior efeito discriminante das amostras de própolis em estudo, haja vista que as quinze ressonâncias de maior significância (Figuras 1.7 e 1.8) ocorrem no intervalo entre 3,0 ppm e 5,50 ppm (Figura 1.2), à semelhança do observado na análise univariada (ANOVA-LSD Fischer). Adicionalmente, o erro de previsão observado nos dois modelos mostrou-se similar, não sendo possível inferir alguma vantagem extra na adoção de uma das técnicas para os dados disponíveis. De fato, considerando-se que as técnicas de RFs e PLS-DA usam formas distintas de inferência e classificação, os seus resultados devem ser interpretados como complementares.

1.5 CONCLUSÕES

A análise metabolômica baseada na determinação dos perfis metabólicos de EHs de própolis catarinenses via espectroscopia uni e bidimensional de RMN, associada a modelos de descrição e classificação, permitiu confirmar a hipótese de trabalho. Dezenove compostos foram identificados nas amostras utilizando-se a 1D/2D-RMN, bases de dados de metabólitos e *softwares* dedicados à análise dos perfis espectrais. De fato, variações importantes de composição química foram detectadas naquela biomassa, com maior significância à ocorrência de compostos com estrutura anomérica, conforme revelado pelos modelos de descrição e classificação adotados. Considerando-se a complexidade química típica de EHs de própolis, a rapidez e a robustez da espectroscopia de RMN e a eficiência das ferramentas de bioinformática descritas para o modelo de estudo, constata-se que a análise metabolômica é uma estratégia relevante a estudos descritivos e de classificação daquela matriz complexa, com importantes efeitos para a aplicação industrial, por exemplo, cosmética e farmacêutica, da própolis, notadamente em sistemas de controle de qualidade.

1.6 PERSPECTIVAS FUTURAS

A metabolômica tem ocupado um lugar importante na bioanálise e na investigação biomédica nos últimos anos, posicionando-se como a mais recente das tecnologias “ômicas”. Um conjunto importante de métodos tem surgido, e novas abordagens analíticas são disponibilizadas a intervalos de tempo que, não raro, demonstram o potencial de desenvolvimento e aplicação dessa área. Este cenário deverá ser ampliado em momentos futuros e de forma rápida, devido à padronização e validação de métodos de coleta, processamento amostral e análise, em paralelo ao contínuo desenvolvimento e aprimoramento do *hardware* e *software* necessários aos estudos metabolômicos. Como exemplo, espectrômetros de RMN equipados com supercondutores de alta potência (isto é, campos magnéticos > 18,8 T – 800 MHz de frequência de ressonância de ^1H) e sondas criogênicas ampliarão significativamente a capacidade de detecção e resolução, otimizando a identificação e a quantificação de metabólitos de menor concentração em matrizes complexas. Concomitantemente, *softwares* dedicados ao processamento e análise do megaconjunto de dados espectrais têm sido desenvolvidos e disponibilizados no mercado a preços acessíveis, quando não gratuitamente. Ademais, o permanente e imprescindível aprimoramento das ferramentas de bioinformática para o tratamento e extração de informações relevantes aos eventos metabólicos claramente sinaliza um futuro de possibilidades concretas quanto à construção de uma visão integrada dos processos bioquímicos, contribuindo para a geração de tecnologias baseadas na genômica funcional nos campos da saúde humana, agricultura e ambiente. Tal perspectiva é ainda mais credível levando-se em conta o desenvolvimento de ferramentas de bioinformática com código aberto e linguagens de bom desempenho, isto é, R, além da disponibilidade em plataformas virtuais na rede mundial de computadores, com livre acesso aos usuários.

REFERÊNCIAS

1. Pirages D. Diversity and social progress in the next millennium: an evolutionary perspective. *Futures*. 2000;32:513-23.
2. Oliver DJ, Nikolau B, Wuertele ES. Functional genomics: high-throughput mRNA, protein, and metabolite analyses. *Metabolic Engineering*. 2002;4:98-106.
3. Costa FF. Non-coding RNAs, epigenetics and complexity. *Gene*. 2008;410:9-17.
4. Sumner LW, Mendes P, Dixon RA. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*. 2003;62:817-36.
5. Fell DA. Beyond genomics. *Trends in Genetics*. 2001;17:680-2.
6. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends in Biotechnology*. 1998;16:373-8.
7. Wood N. Profiling modified metabolomes. *Trends in Plant Science*. 2001;6:191-3.
8. Fernie AR, Schauer N. Metabolomics-assisted breeding: a viable option for crop improvement? *Trends in Genetics*. 2008;25:39-48.
9. Keurentjes JJB. Genetical metabolomics: closing in on phenotypes. *Current Opinion in Plant Biology*. 2009;12:223-30.
10. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Ressler-Tunali U, Beaçe MH, Trethewey RN, Lange BM, Wuertele ES, Sumner LW. Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*. 2004;9:418-25.
11. Goodcare R, Vaidynathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in Biotechnology*. 2004;22:245-52.
12. Fukusaki E, Kobayashi A. Plant metabolomics: potential for practical operation. *Journal of Bioscience and Bioengineering*. 2005;100:347-54.
13. Saghatelian A, Cravatt BF. Global strategies to integrate the proteome and metabolome. *Current Opinion in Chemical Biology*. 2005;9:62-8.
14. Abhyankar G, Reddy VD, Giri CC, Rao KV, Lakshimi, VVS, Prabhakar S, Vairamani M, Thippeswamy BS, Bhattacharya PS. Amplified fragment length polymorphism and metabolomic profiles of hairy roots of *Psoralea corylifolia* L. *Phytochemistry*. 2005;66:2441-57.
15. Mitchell SC, Carmichael PL. Metabonomics and the endocrine system. *Molecular and Cellular Endocrinology*. 2005;244:10-4.
16. Wind RA, Hu J, Mayors PD. Slow-MAS NMR: A new technology for in vivo metabolomic studies. *Drug Discovery Today: Technologies*. 2005;2:283-86.
17. Tanaka R, Nitta A, Nagatsu A. Application of a quantitative ^1H -NMR method for the determination of amygdalin in *Persicae* sêmen, *Armeniaceae* sêmen, and *Mume fructus*. *Journal of Natural Medicines*. 2014;68:225-30.

18. Kuhn S, Lemos PMM, Oliveira SK, Silva DA, Tomazzoli MM, Souza ACV, Pinto MRR, Uarrota VG, Cella I, Ferreira AG, Zeggio AS, Veleirinho MBR, Delgadillo I, Vieira FN, Maraschin M. Metabolomics and chemometrics as tools for chemo(bio)diversity analysis – maize landraces and propolis. In: Chemometrics in practical applications. Rijeka: InTech; 2012. Capítulo 1; p. 253-70.
19. Leardi R. Chemometrics in data analysis. In: A user-friendly guide to multivariate calibration and classification. West Sussex: NIR Publications; 2003.
20. Verpoorte R, Maraschin M. Aplicações da cromatografia líquida e espectrometria de massa na análise de metabólitos secundários vegetais e em biomedicina. In: Plantas Medicinais sob a Ótica da Química Medicinal Moderna. 1 ed. Chapecó: Unoesc; 2001. Capítulo 1; p. 147-93.
21. Stegehuis DS. On-line chromatographic and electrophoretic techniques for sample pretreatment in bioanalysis [PhD thesis]. Leiden: Leiden University; 1992. 188p.
22. Allwood JW, Erban A, Koning S, Dunn WB, Luedemann A, Lommen A, Kay L, Loscher R, Kopka L, Goodacre R. Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC–EI–TOF/MS) based plant metabolomics. *Metabolomics*. 2009;5:479-96.
23. Horning EC, Horning MG. Human metabolic profiles obtained by GC and GC/MS. *Journal of Chromatographic Science*. 1971a;9:129-40.
24. Horning EC, Horning MG. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clinical Chemistry*. 1971b;17:802-9.
25. Gates SC, Sweeley CC. Quantitative metabolic profiling based on gas chromatography. *Clinical Chemistry*. 1978;24(10):1663-73.
26. Bartle KD, Myers P. History of gas chromatography. *Trends in Analytical Chemistry*. 2002;21:547-57.
27. Nishiumi S, Kobayashi T, Ikeda A, Yoshie T, Kibi M, Izumi Y, Okuno T, Hayashi N, Kawano S, Takenawa T, Azuma T, Yoshida M. A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS One*. 2012;7:e40459.
28. Tinke AP, van der Hoeven RAM, Niessen WMA, Tjaden UR, van der Greef J. Some aspects of peak broadening in particle-beam liquid chromatography-mass spectrometry. *Journal of Chromatography*. 1991;554:119-24.
29. M'Koma AE, Blum DL, Norris JL, Koyama T, Billheimer D, Motley S, Ghiassi M, Ferdowsi N, Bhowmick I, Chang SS, Fowke JH, Caprioli RM, Bhowmick NA. Detection of pre-neoplastic and neoplastic prostate disease by MALDI profiling of urine. *Biochemical and Biophysical Research Communications*. 2007;353:829-34.
30. Bertram HC, Eggers N, Eller N. Potential of human saliva for nuclear magnetic resonance-based metabolomics and for health-related biomarker identification. *Analytical Chemistry*. 2009;81:9188-93.

31. Chassy BM. Can-omics inform a food safety assessment? *Regulatory Toxicology and Pharmacology*. 2010;58:S62-70.
32. Doerrer N, Ladics G, McClain S, Herouet-Guicheney C, Poulsen LK, Privalle L, Stagg N. Evaluating biological variation in non-transgenic crops: executive summary from the ILSI Health and Environmental Sciences Institute workshop, November 16-17, 2009, Paris, France. *Regulatory Toxicology and Pharmacology*. 2010;58:S2-7.
33. Madsen R, Lundstedt T, Trygg T. Chemometrics in metabolomics – A review in human disease diagnosis. *Analytica Chimica Acta*. 2010;659:23-33.
34. Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical Biochemistry*. 2007;363:185-95.
35. Gomez-Casati DF, Zanol MI, Busi MV. Metabolomics in plants and humans: applications in the prevention and diagnosis of diseases. *BioMedical Research International*. 2013;Article ID 792527.
36. Kawakami M, Kobayashi A. Volatile constituents of green mate and roasted mate. *Journal of Agricultural and Food Chemistry*. 1991;39:1275-9.
37. Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AWM, Verpoorte R. Metabolomic differentiation of *Cannabis sativa* cultivars using ¹H NMR spectroscopy and principal component analysis. *Journal of Natural Products*. 2004;67:953-7.
38. Yuliana ND, Khatib A, Choi YH, Verpoorte R. Metabolomics for bioactivity assessment of natural products. *Phytotherapy Research*. 2011;25:157-69.
39. Davies HV, Shepherd LVT, Stewart D, Frank T, Röhlig RM, Engel KH. Metabolome variability in crop plant species – When, where, how much and so what? *Regulatory Toxicology and Pharmacology*. 2010;58:S54-61.
40. Kuhn S, Ogliari JB, Dias PF, Santos MS, Ferreira AG, Bonham CC, Wood KV, Maraschin M. Metabolic fingerprint of Brazilian maize landraces silk (stigma/styles) using NMR spectroscopy and chemometric methods. *Journal of Agricultural and Food Chemistry*. 2010a;58:2194-200.
41. Kuhn S, Ogliari JB, Dias PF, Boffo EF, Correia I, Ferreira AG, Delgadillo I, Maraschin M. ATR-FTIR spectroscopy and chemometric analysis applied to discrimination of landrace maize flours produced in southern Brazil. *International Journal of Food Science & Technology*. 2010b;45:1673-81.
42. Marcucci MC. Propriedades biológicas e terapêuticas dos constituintes químicos da própolis. *Química Nova*. 1996;19:529-36.
43. Fontana JD, Adelman J, Passos M, Maraschin M, Lacerda CA, Lanças FM. Propolis: chemical micro-heterogeneity and bioactivity. *Methods in Biotechnology*. 2004;16:203-18.
44. Pereira AS, Seixas FRMS, Aquino FRA. Própolis: 100 anos de pesquisa e suas perspectivas futuras. *Química Nova*. 2002;25:321-6.
45. Castaldo S, Capasso F. Propolis, an old remedy used in modern medicine. *Fitoterapia*. 2002;73:S1-6.

46. Woisky RG, Giesbrecht AM, Salatino A. Atividade antibacteriana de uma formulação preparada a partir de própolis de *Apis mellifera* L. *Revista de Farmácia e Bioquímica da Universidade de São Paulo*. 1994;30:19-21.
47. Lustosa SR. Própolis: atualizações sobre a química e a farmacologia. *Revista Brasileira de Farmacognosia*. 2008;18:447-54.
48. Salatino A, Teixeira EW, Negri G, Message D. Origin and chemical variation of Brazilian propolis. *Evidence Based Complementary and Alternative Medicine*. 2005;2:33-8.
49. Ghisalberti EL. Propolis: a review. *Bee World*. 1979;12:59-84.
50. Bankova V. Chemical diversity of propolis and the problem of standardization. *Journal of Ethnopharmacology*. 2005;100:114-77.
51. Souza EA, Inoue HT, Gomes SMA, Funari SRC, Orsi RO. Propriedade físico-química da própolis em função da sazonalidade e método de produção. *Archivos de Zootecnia*. 2010;59:210-4.
52. Park YK, Alencar SM, Aguiar CL. Botanical origin and chemical composition of Brazilian propolis. *Journal of Agricultural and Food Chemistry*. 2002;50:2502-6.
53. Koo H, Gomes BPFA, Rosalen PL, Ambrosano GMB, Park YK, Cury JA. In vitro antimicrobial activity of propolis and *Arnica montana* against oral pathogens. *Archives of Oral Biology*. 2006;45:141-8.
54. Duarte S, Rosalen PL, Hayacibara MF, Cury JA, Rehder VG, Sartorato A, Ikegami M, Bowen WH, Koo H. The influence of a novel propolis on mutans streptococci biofilms and caries development in rats. *Caries Research*. 2005;39:328-9.
55. Marcucci MC. Própolis tipificada: um novo caminho à elaboração de medicamentos de origem natural contendo este produto apícola. *Revista Fitos*. 2006;1:36-46.
56. Bankova V, Marcucci MC, Simova S, Nikolova N, Kujumgiev A, Popov S. Antibacterial diterpenic acids from Brazilian propolis. *Zeitschrift fur Naturforschung C*. 1996;50:277-80.
57. Bankova V, Christov R, Kujumgiev A, Marcucci MC, Popov S. Chemical composition and anti-bacterial activity of Brazilian propolis. *Zeitschrift fur Naturforschung C*. 1995;50:167-72.
58. Banskota AH, Tezuka Y, Prasain JK, Matsushige K, Saiki I, Kadota S. Chemical constituents of Brazilian propolis and their cytotoxic activities. *Journal of Natural Products*. 1998;7:896-900.
59. Alencar SM, Aguiar CL, Paredes-Guzmán J, Park YK. Composição química de *Baccharis dracunculifolia*, fonte botânica das própolis dos estados de São Paulo e Minas Gerais. *Ciência Rural*. 2005;35:909-15.
60. Soares AKA, Carmo GC, Quental DP, Nascimento DF, Bezerra FAF, Moraes MO, Moraes MEA. Clinical safety evaluation of a phytomedicine containing *Mikania glomerata*, *Grindelia robusta*, *Copaifera officinalis*, *Myroxylon toluifera*, *Nasturtium*

officinale, honey and propolis in healthy volunteers. *Revista Brasileira de Farmacognosia*. 2006;16:447-54.

61. Tavares JP, Martins IL, Vieira AS, Lima FAV, Bezerra FAF, Moraes MO, Moraes MEA. Estudo de toxicologia clínica de um fitoterápico a base de associações de plantas, mel e própolis. *Revista Brasileira de Farmacognosia*. 2006;16:350-6.

62. Bankova V. Chemical diversity of propolis is a valuable source of new biologically active compounds. *Journal of ApiProducts and ApiMedical Science*. 2009;1:23-8.

63. Grundberger D, Banerjee R, Eisinger K, Oltz EM, Efros L, Caldwell M, Estevez V, Nakanishi K. Preferential cytotoxicity on tumor cell by caffeic acid phenethyl ester isolated from propolis. *Experientia*. 1988;44:230-2.

64. Velazquez C, Navarro M, Acosta A, Angulo A, Dominguez Z, Robles R, Robles-Zepeda R, Lugo E, Goycoolea FM, Velazquez EF, Astiazaran H, Hernandez J. Antibacterial and free-radical scavenging activities of Sonoran propolis. *Journal of Applied Microbiology*. 2007;103:1747-56.

65. Pommier Y, Johnson AA, Marchand C. Integrase inhibitors to treat HIV/AIDS. *Nature Review Drug Discovery*. 2005;4:236-48.

66. Popova M, Bankova V, Butovska D, Petkov V, Damyanova B, Sabatini AG, Marcuzzan GL, Bogdanov S. Poplar type propolis and analysis of its biologically active components. *Honeybee Science*. 2003;24:61-6.

67. Popova M, Bankova V, Butovska D, Petkov V, Nikolova-Damyanova B, Sabatini AG, Marcuzzan GL, Bogdanov S. Validated methods for the quantification of biologically active constituents of poplar-type propolis. *Phytochemical Analysis*. 2004;15:235-40.

68. Boudorova-Kravsteva G, Bankova V, Sforcin JM, Nikolova N, Popov S. Phenolics from Brazilian propolis. *Zeitschrift fur Naturforschung C*. 1997;52:676-9.

69. Marcucci MC, Ferreres F, Custodio AR, Ferreira MM, Bankova VS, Garcia-Viguera C, Bretz WA. Evaluation of phenolic compounds in Brazilian propolis from different geographic regions. *Zeitschrift fur Naturforschung C*. 2000;55:76-81.

70. Park YK, Ikegaki M, Alencar SM, Moura FF. Evaluation of Brazilian propolis by both physicochemical methods and biological activity. *Honeybee Science*. 2000;21:85-90.

71. Bankova V, Castro SL, Marcucci MC. Propolis: recent advances in chemistry and plant origin. *Apidologie*. 2000;31:3-15.

72. Marcucci MC, Ferreres F, Garcia-Viguera C, Bankova VS, Castro SL, Dantas AP, Valente PH, Paulino N. Phenolic compounds from Brazilian propolis with pharmacological activities. *Journal of Ethnopharmacology*. 2001;74:105-12.

73. Watson DG, Peyfoon E, Zheng L, Lu D, Seidel V, Johnston B, Parkinson JA, Fearnley J. Application of principal components analysis to ¹H-NMR data obtained from propolis samples of different geographical origin. *Phytochemical Analysis*. 2006;17:323-31.

74. Cuesta-Rubio O, Piccinelli AL, Fernandez MC, Hernandez IM, Rosado A, Rastrelli L. Chemical characterization of Cuban propolis by HPLC-PDA, HPLC-MS, and NMR:

the brown, red, and yellow Cuban varieties of propolis. *Journal of Agricultural and Food Chemistry*. 2007;55:7502-9.

75. Maraschin M, Somensi-Zeggio A, Oliveira SK, Kuhnen S, Tomazzoli MM, Zeri ACM, Carreira R, Rocha M. A machine learning and chemometrics assisted interpretation of spectroscopic data – A NMR-based metabolomics platform for the assessment of Brazilian propolis. *Lecture Notes in Computer Science*. 2012;7632:129-40.

76. Papotti G, Bertelli D, Plessi M, Rossi MC. Use of HR-NMR to classify propolis obtained using different harvesting methods. *International Journal of Food Science and Technology*. 2010;45:1610-8.

77. Maraschin RP, Ianssen C, Ogliari JB, Dias PF, Maraschin M, Cimadon AMA, Zanús C, Caro MSB. Flash-chromatography and ¹H-NMR analysis of Brazilian Cabernet Sauvignon wines – a chemical composition similarity study. *Annals of Magnetic Resonance*. 2003;2:48-51.

78. Kim HK, Choi YH, Verpoorte R. NMR-based metabolomic analysis of plants. *Nature Protocols*. 2010;5:536-49.

79. Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*. 2009;37:W652-60.

80. Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*. 1901;2:559-72.

81. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna: The R Foundation for Statistical Computing; 2013 [Internet]. [Cited 2016 August 11]. Available from: <http://www.R-project.org>.

82. Johnson SC. Hierarchical clustering schemes. *Psychometrika*. 1967;32(3):241-54.

83. Barker M, Rayens W. Partial least squares for discrimination. *Journal of Chemometrics*. 2003;17:166-73.

84. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, van Duijnhoven JPM, van Dorsten FA. Assessment of PLSDA cross-validation. *Metabolomics*. 2008;4:81-9.

85. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006;7:142-7.

86. Witten H, Frank E, Hall MA. Data Mining – Practical Machine Learning Tools and Techniques. 3rd ed. San Francisco: Morgan Kaufman; 2011.

87. Weljie A, Newton J, Mercier P, Carlson E, Slupsky C. Targeted profiling: quantitative analysis of ¹H-NMR metabolomics data. *Analytical Chemistry*. 2006;78:4430-42.